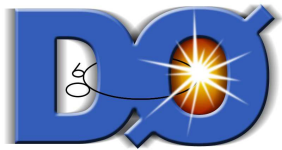
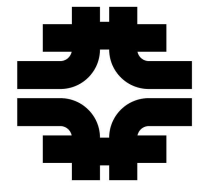


Remote Computing



Daniel Wicke
(Fermilab)

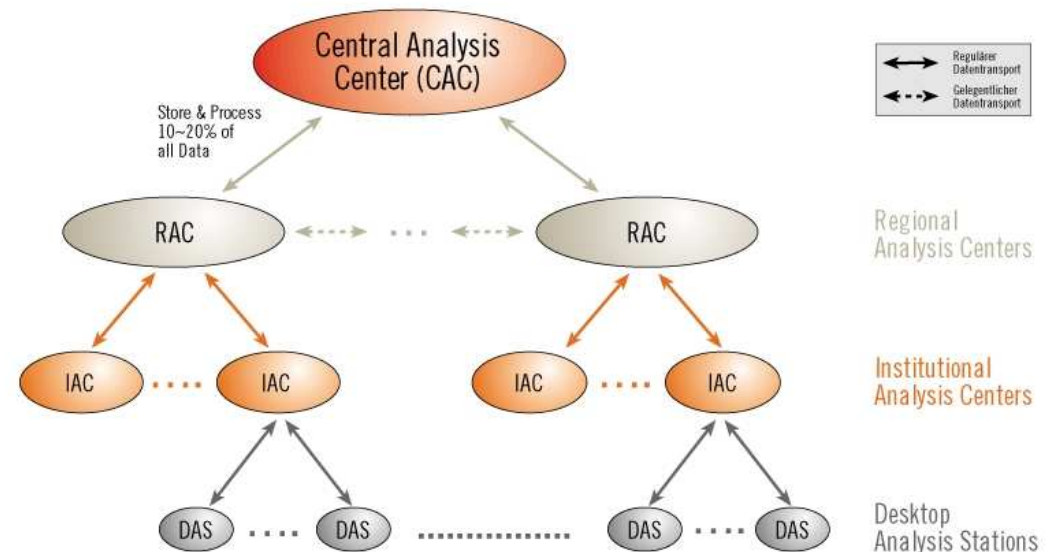


Outline

- Introduction
- Monte Carlo Production
- Data Reprocessing
- Summary

Introduction

- Remote computing has been in DØ's plan since ~ 1997 .
 - All Monte Carlo for RunII has been produced remotely.
 - SAM to be used for data handling.
- Since 2002 DØ is increasing its offsite computing usage:
 - Regional Analysis Centers established a tiered structure for data access.
 - Allows (manual) remote production and analysis
- Now moving towards GRID
 - Monte Carlo
 - Data Processingwith unified/centralised submission.



Monte Carlo Production

- Over the last year DØ produced around 160M Monte Carlo events (~ 7 TB).
- These were produced at 10 different remote sites:

Resources

IN2P3, Lyon	local
Nikhef	local/LCG
Tata (SAR)	local/mcfarm
UTA (SAR)	local/mcfarm
Sprace (SAR)	local/mcfarm
Ouhep (SAR)	SAMGrid/mcfarm
Luhep (SAR)	SAMGrid/mcfarm
LTU (SAR)	SAMGrid/mcfarm
Prague	SAMGrid
GridKa	SAMGrid

- Submission is person power intense
- Unified system helps to take advantage from improvements at many sites
 \Rightarrow **mcfarm**
- Gridified system helps to reduce the number of required operators.
 \Rightarrow **SAMGrid**

Data Reprocessing

Motivation: Improved understanding of the DØ-Detector

- We have improved calorimeter calibration in p17
- Basis:
 - improved understanding of the detector
 - based on reality rather than design/plans
- All of our data were reconstructed with p14

⇒ *Redo reconstruction of all data*

The Computing Task

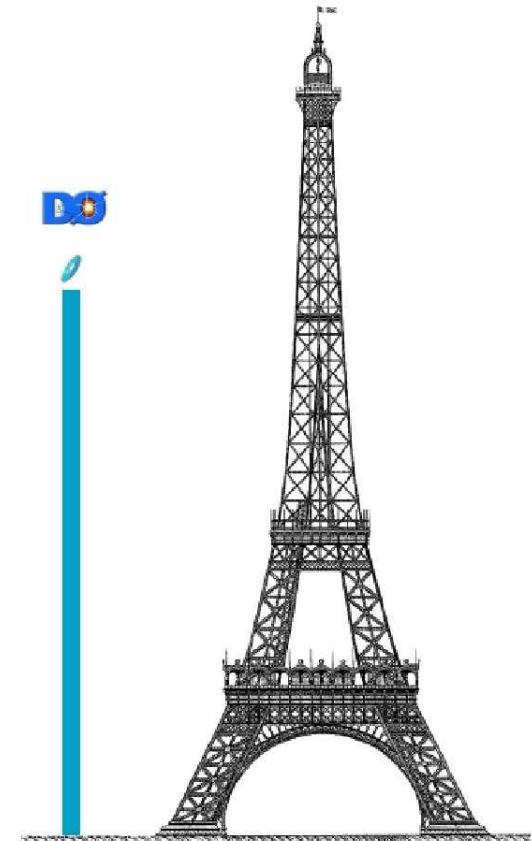
	p17 reprocessing	p14 reprocessing
Luminosity	470 pb ⁻¹	100 pb ⁻¹
Events	1G	300M
Rawdata 250kB/Event	250TB	75TB
DSTs 150kB/Event	150TB	45TB
TMBs 70kB/Event	70TB	6TB
Time 50s/Event	20,000months	6000months
(on 1GHz Pentium III)	3400CPUs for 6mths	2000CPUs for 3mths
Remote processing	100%	30%

Central Farm (1000CPUs) used to capacity with data taking.

The Computing Task

	p17 reprocessing
Luminosity	470 pb^{-1}
Events	1G
Rawdata 250kB/Event	250TB
DSTs 150kB/Event	150TB
TMBs 70kB/Event	70TB
Time 50s/Event	20,000months
(on 1GHz Pentium III)	3400CPUs for 6mths
Remote processing	100%

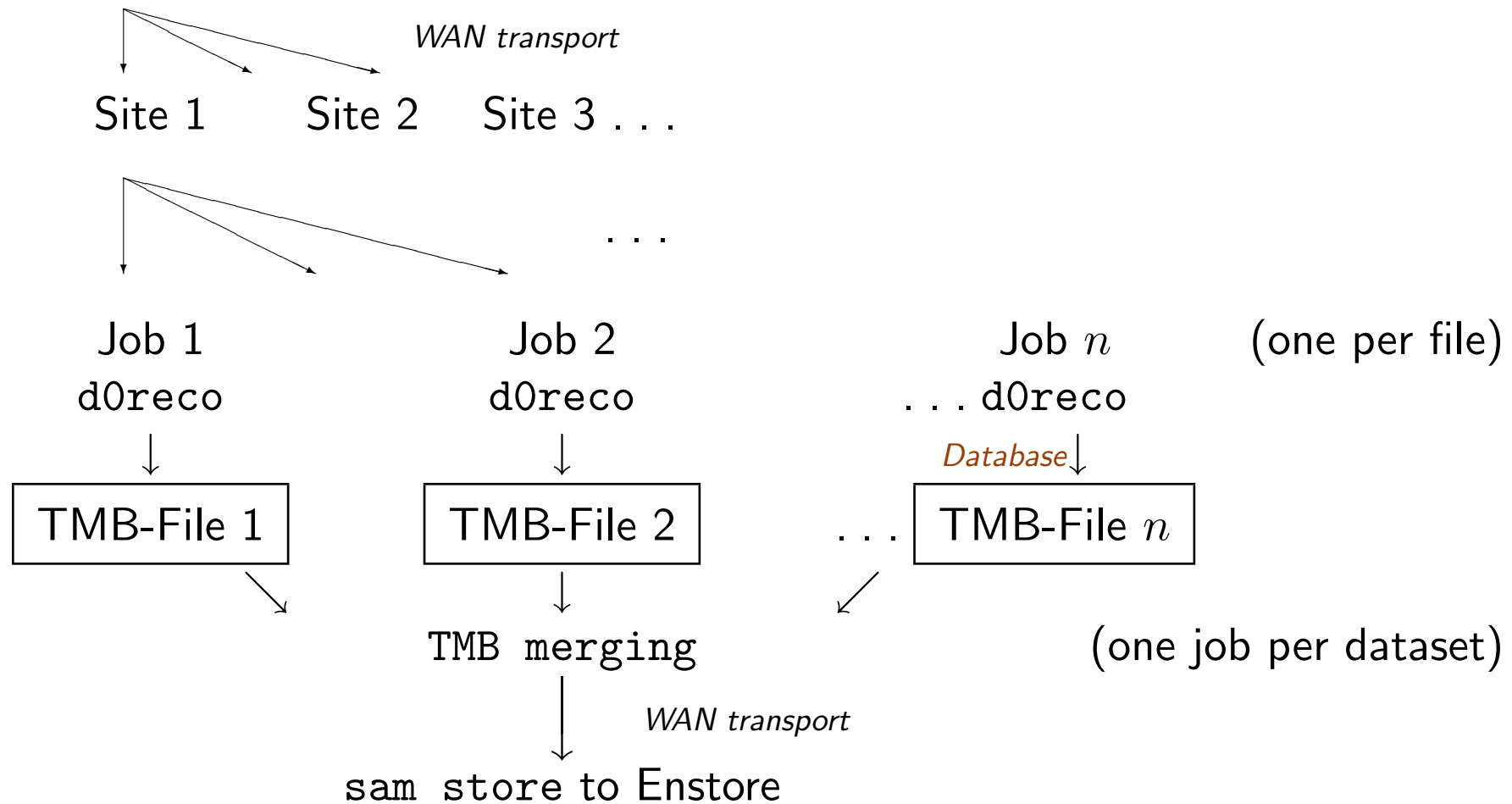
A stack of CDs as high as the Eiffeltower



Application flow

Overview

Datasets of RAW-files



Implementation

SAMGrid was chosen to implement this task on distributed systems.

- Each dataset processed through d0reco in one grid job.
- The corresponding merge job submitted separately .

Using a grid ...

- provides common environment for d0reco at all sites.
- allows common operation scripts (d0repro).
 - submission (and recovery) is done by
sub_production.py <dataset> <d0release>
sub_merge.py <dataset> <d0release>
 - production and merge status can be checked (poor man's request system)

Tests on the 700CPU DØFarm revealed scalability issues in JIM

Behaviour was improved by a factor of 60(!).

Error Handling and Recovery

Beside unrecoverable crashes of d0reco there will be *random* crashes.

(Network outages, file delivery failures, batch system crashes/hangups, worker-node crashes, filesystem corruption...)

Book-keeping

1. of succeeded jobs/files

needed to assure completion without duplicated events.

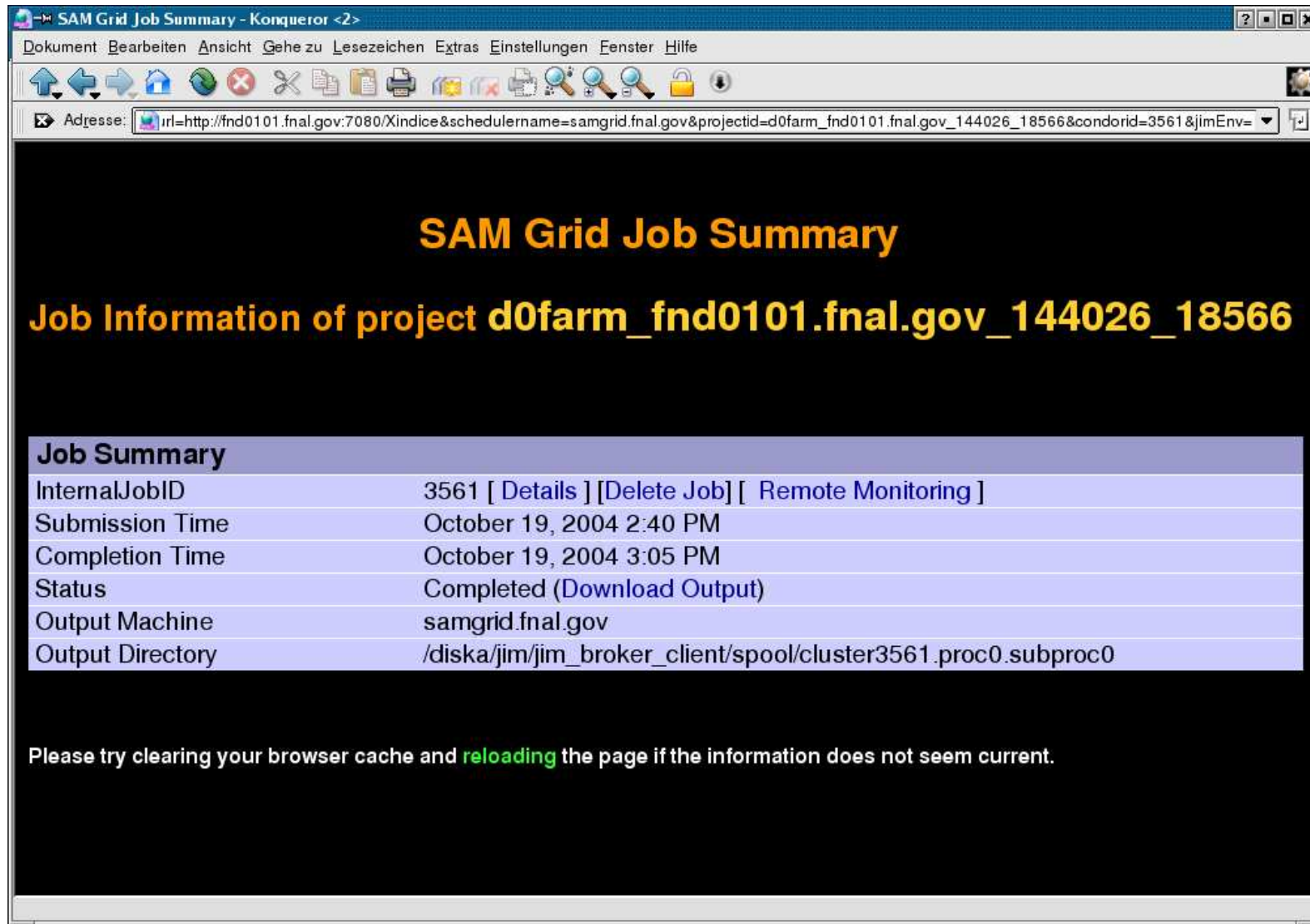
SAM is used avoid data duplication and to define recovery jobs.

2. of failed jobs/files

needed to trace problems in order fix bugs and to assure efficiency.

JIMs XML-DB is used to ease bug tracing and provide fast recovery.

Some Screen-shots



SAM Grid Job Summary

Job Information of project d0farm_fnd0101.fnal.gov_144026_18566

Job Summary	
InternalJobID	3561 [Details] [Delete Job] [Remote Monitoring]
Submission Time	October 19, 2004 2:40 PM
Completion Time	October 19, 2004 3:05 PM
Status	Completed (Download Output)
Output Machine	samgrid.fnal.gov
Output Directory	/diska/jim/jim_broker_client/spool/cluster3561.proc0.subproc0

Please try clearing your browser cache and **reloading** the page if the information does not seem current.

Some Screen-shots (2)

Sam Grid RunJob Details - Konqueror

Location: arm_fnd0101.fnal.gov_181510_10259&Dburl=http://fnd0101.fnal.gov:7080/Xindice&schedulename=samgrid.fnal.gov&condorid=5085&requestId=&numevents=&jobType=dzero_reconstruction

Sam Grid RunJob Details

Cluster Id (Sandbox no) = 23353.1110421577

Reconstruction DataSet Id = dayset-2004-06-22-194374-0

To list files produced by this job select data_tier and click the button

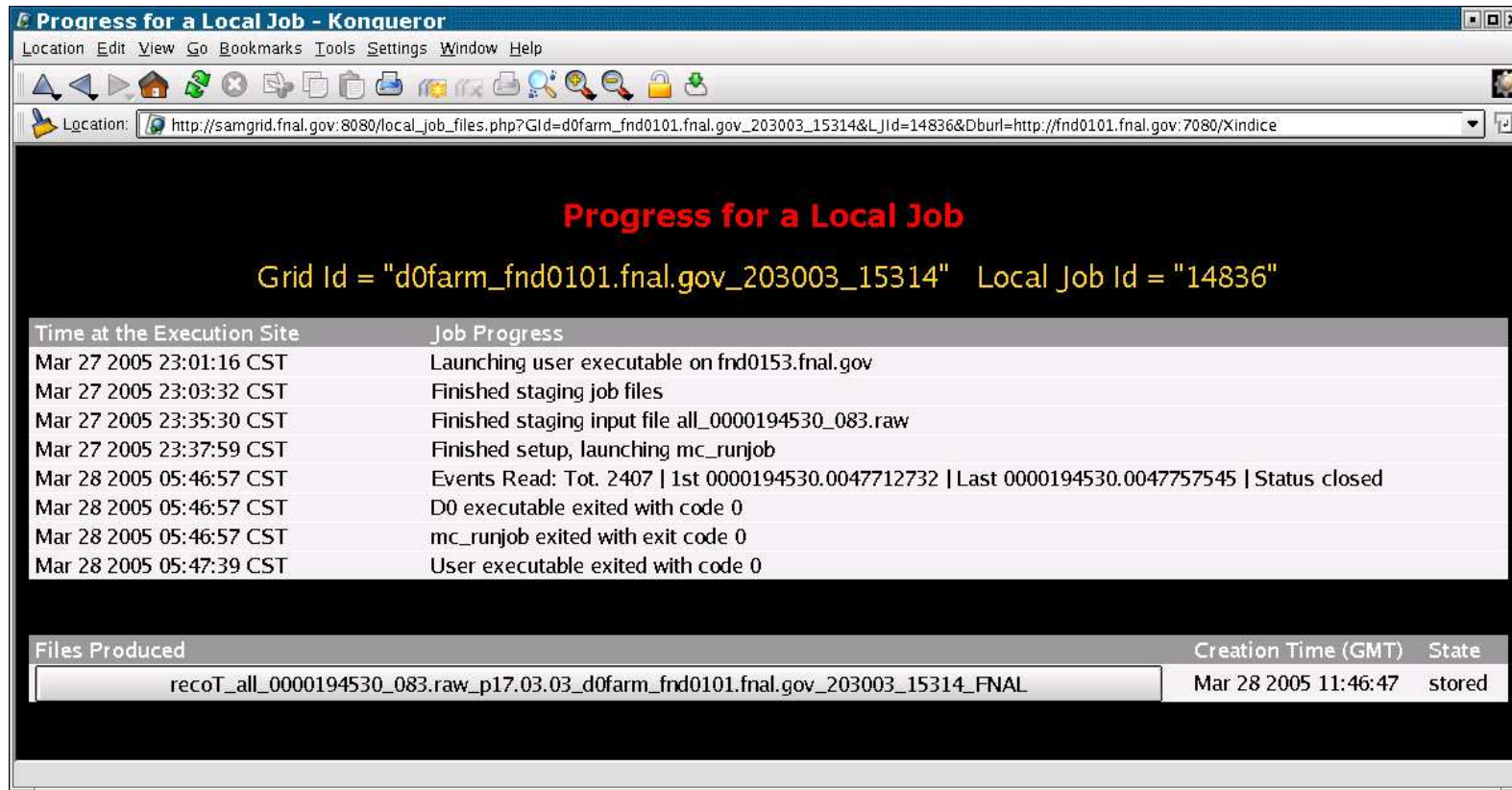
All Files Query SAM

[List files known to xmladb](#)

Grid Id	Created (GMT)	Total jobs	Finished	Running	Queued
d0farm_fnd0101.fnal.gov_181510_10259	Mar 09 2005, 20:31:13 CST	99	98	1	0

No	Local Job Id	Status	Time Stamp (GMT)	Check Progress	No of Outputfiles	Delete
1	13297	done (User executable exited with code 0)	Mar 10 2005, 06:50:46 CST	Monitor Progress	1	[X]
2	13298	done (User executable exited with code 0)	Mar 10 2005, 07:11:16 CST	Monitor Progress	1	[X]
3	13299	active (User executable exited with code 0)	Mar 09 2005, 20:31:24 CST	Monitor Progress	0	[X]
4	13300	done (User executable exited with code 0)	Mar 10 2005, 07:37:01 CST	Monitor Progress	1	[X]
5	13301	done (User executable exited with code 0)	Mar 10 2005, 06:20:03 CST	Monitor Progress	1	[X]
6	13302	done (User executable exited with code 0)	Mar 10 2005, 07:26:46 CST	Monitor Progress	1	[X]
7	13303	done (User executable exited with code 0)	Mar 10 2005, 09:23:55 CST	Monitor Progress	1	[X]
8	13304	done (User executable exited with code 0)	Mar 10 2005, 08:12:50 CST	Monitor Progress	1	[X]
9	13305	done (User executable exited with code 0)	Mar 10 2005, 09:03:40 CST	Monitor Progress	1	[X]

Some Screen-shots (3)



The screenshot shows a web browser window titled "Progress for a Local Job - Konqueror". The address bar displays the URL: http://samgrid.fnal.gov:8080/local_job_files.php?GId=d0farm_fnd0101.fnal.gov_203003_15314&LJId=14836&Dburl=http://fnd0101.fnal.gov:7080/Xindice.

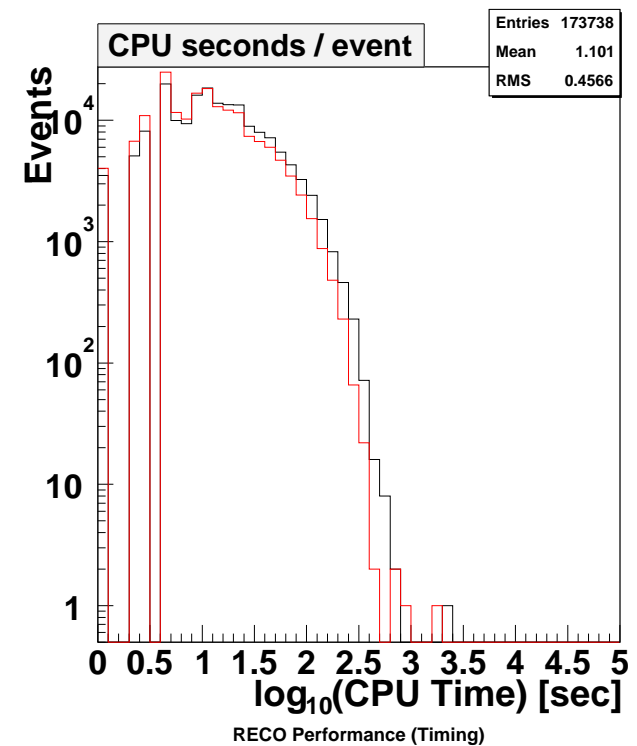
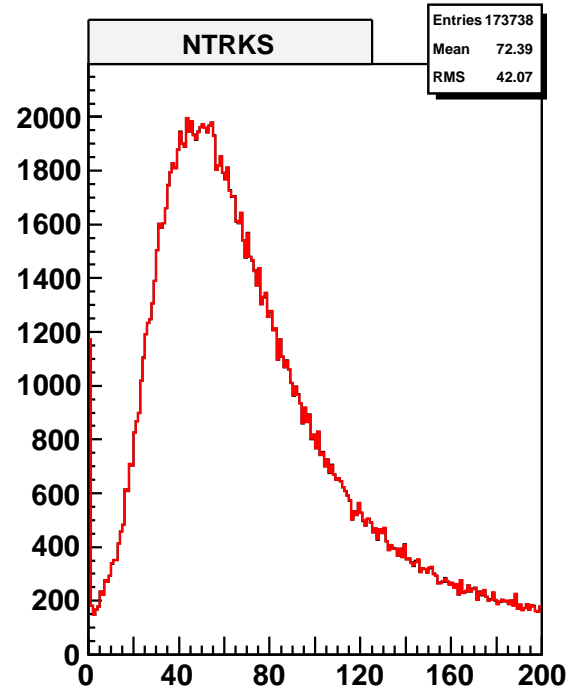
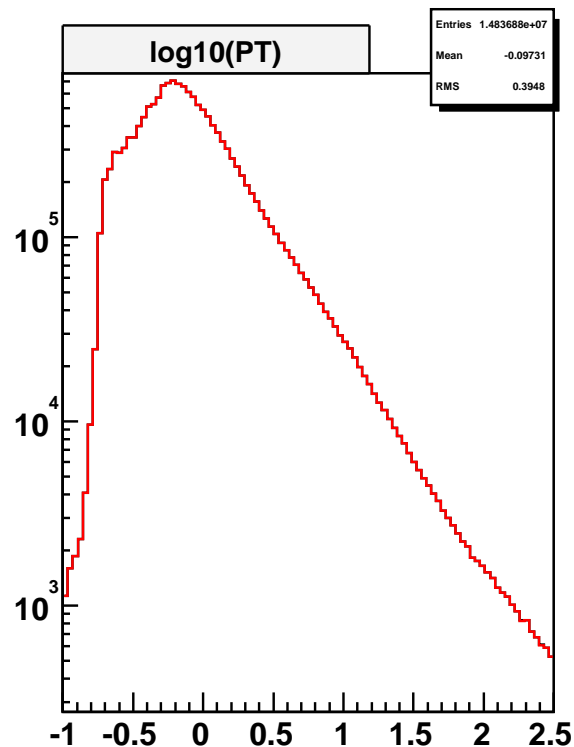
The main content area has a black background with red text "Progress for a Local Job" and yellow text "Grid Id = 'd0farm_fnd0101.fnal.gov_203003_15314' Local Job Id = '14836'".

Time at the Execution Site	Job Progress
Mar 27 2005 23:01:16 CST	Launching user executable on fnd0153.fnal.gov
Mar 27 2005 23:03:32 CST	Finished staging job files
Mar 27 2005 23:35:30 CST	Finished staging input file all_0000194530_083.raw
Mar 27 2005 23:37:59 CST	Finished setup, launching mc_runjob
Mar 28 2005 05:46:57 CST	Events Read: Tot. 2407 1st 0000194530.0047712732 Last 0000194530.0047757545 Status closed
Mar 28 2005 05:46:57 CST	D0 executable exited with code 0
Mar 28 2005 05:46:57 CST	mc_runjob exited with exit code 0
Mar 28 2005 05:47:39 CST	User executable exited with code 0

Files Produced	Creation Time (GMT)	State
recoT_all_0000194530_083.raw_p17.03.03_d0farm_fnd0101.fnal.gov_203003_15314_FNAL	Mar 28 2005 11:46:47	stored

Certification of Sites

- Each center needs to process agreed datasets (100 files) for certification.
- Unmerged and merged TMBs are compared per site.
- Common set of events is compared between sites.



Available Resources

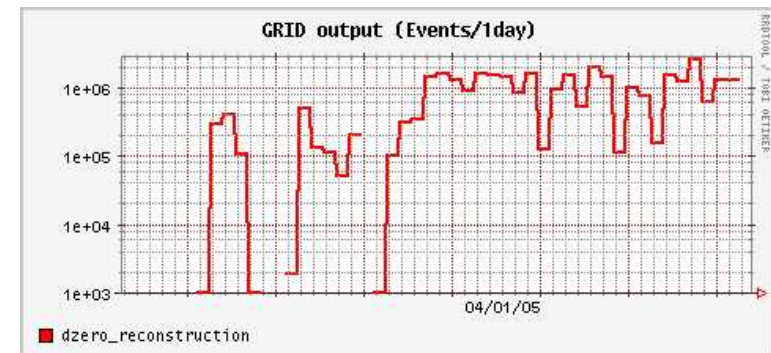
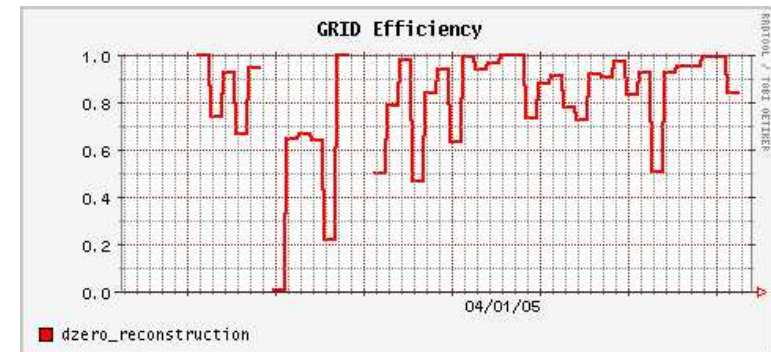
FNAL Farm,	1000CPUs	SAMGrid	used by data-taking
Westgrid,	600CPUs	SAMGrid	running
Lyon,	400CPUs	SAMGrid	running
Wisconsin,	30CPUs	SAMGrid	certified
Prague,	200CPUs	SAMGrid	certified
SAR (UTA),	230CPUs	SAMGrid	certifying
GridKa,	500CPUs	SAMGrid	certifying
CMS Farm,	100CPUs	LCG with JIM jobmanger	under test
UK (4 sites)	750CPUs	SAMGrid	2 certifying
<hr/>			
External	~2800CPUs	(1GHz PIII equiv.)	

Running 35%, About to start 35% , To join soon 30%.

⇒ *not* sufficient for completion in 6mths, but d0reco is faster than projected

Outlook

- Production
 - Work on efficiency.
 - Install more sites if available.
Oscer, SPrace volunteering
- SAM Grid
 - add brokering
⇒ decrease person power required
 - interface SAMGrid to LCG
⇒ increase CPU resources
- Operation scripts
 - auto pilot in d0repro
⇒ decrease person power required



Summary

- Monte Carlo Production
 - gradually migrating from distributed operation to common tools
 - further to a gridified operation.
 - done during full load production.
- Data reprocessing effort more than $3\times$ bigger than the 2003/4 effort.
 - 250TB to be distributed.
 - Est. 1600CPU years to produce 70TB.
 - Fully gridified.
- Fermilab is working on a press release:
Fermilab's DZero Experiment Crunches Record Data with the Grid

